



## Deliverable 3.2

# Publish formalised EuroFAANG Data policy and access principles

**Grant agreement no°: 101094718**

Due submission date

**2023-12-31**

Actual submission date

**2024-01-15**

Responsible author(s)

**Peter Harrison**

**peter@ebi.ac.uk**

**Confidential: No**

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094718. The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

## DOCUMENT CONTROL SHEET

Deliverable name	Publish formalised EuroFAANG Data policy and access principles
Deliverable number	D3.2
Partners providing input to this Deliverable	EMBL, WU, INRAE, FBN
Draft final version circulated by lead party to: On date	Coordinators and Work Package Leaders on 2024-01-08
Approved by (on date)	Coordinators on 2024-01-15
Work package no	3
Dissemination level	Public (PU)

## REVISION HISTORY

Version number	Version date	Document name	Lead partner
V1	2024-01-08	D3.2_EuroFAANG_Data_policy_and_access_principles_V1	EMBL
V2	2024-01-12	D3.2_EuroFAANG_Data_policy_and_access_principles_V2	EMBL

### Changes with respect to the DoA (Description of Action)

None

### Dissemination and uptake

This is a public deliverable.

## Table of Content

<b>Changes with respect to the DoA (Description of Action)</b> .....	<b>1</b>
<b>Dissemination and uptake</b> .....	<b>1</b>
<b>1. Introduction</b> .....	<b>3</b>
<b>2. EuroFAANG Data Policy and Access Principles Document</b> .....	<b>3</b>
1. The EuroFAANG RI .....	5
2. EuroFAANG Open access policy.....	5
3. FAANG Data Sharing Policy.....	6
4. FAIR data policy .....	10
5. Long term data management and preservation of data .....	11
6. Brokered submission requirements and metadata standards.....	11
7. Access procedures .....	12
8. Data Re-use and Citation .....	12
9. Role of EuroFAANG Data Coordination Centre .....	12
10. Monitoring and Review .....	12
11. Contact and Support.....	12
12. Additional Considerations .....	13
<b>3. Conclusions</b> .....	<b>14</b>

## **1. Introduction**

The purpose of this deliverable is to report on the establishment and publishing of the first version of a formalised EuroFAANG Data policy and access principles. The policy and access principles set out for the emerging infrastructure expectations for how future users' data will be handled and subsequently accessed. The policy was developed to mirror and accompany the EuroFAANG Transnational Access policy, that was previously reported in deliverable 2.1. This Transnational Access policy was additionally supplemented by a document outlining governance policy on different options for transnational access, as reported in deliverable 2.2. The Transnational Access and Data Access policies are designed to be viewed in unison as a developing guide for future users that wish to access the infrastructure services and/or gain access to the infrastructures generated genomic and phenotypic data.

This EuroFAANG Data policy and access principles outlines the EuroFAANG Research Infrastructures focus on open science and FAIR data principles. The policies have been developed in consultation and in the context of existing global FAANG data principles, ELIXIR standards and existing European policies. This policy follows the publishing last year of the EuroFAANG Research Infrastructures Data Management Plan, that was developed in accordance with EU Horizon guidelines, and set out the key focus of the infrastructure on open science and FAIR principles.

Just like the Data Management Plan and the Transnational Access Policy, this Data policy will be a living document that will be updated regularly throughout the lifecycle of the research infrastructure to reflect the evolving needs of the infrastructure and the communities it supports. The data policy and access principles will also develop based on the EuroFAANG work package and think tank outcomes around data generation, processing, and transnational access, that in particular focus on genome editing, biobanks, phenotype recording, and the emerging interactions with other key EU Research Infrastructures and services.

This policy and access principles are designed to ensure open and FAIR access to high quality data to ensure exploitation of results and enable effective genotype to phenotype research in farmed animals by the European research community and beyond.

## **2. EuroFAANG Data Policy and Access Principles Document**



## EuroFAANG Data Policy and Access Principles

Version: 2.0

Published: 15<sup>th</sup> January 2024

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094718. The content of this document reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

The EuroFAANG Research Infrastructure (referred to as "EuroFAANG RI" hereinafter) is committed to open science principles and recognizes the critical role of research data in advancing knowledge and fostering innovation. This data policy document outlines the principles and procedures governing access to and reuse of data generated by and in collaboration with the EuroFAANG RI in accordance with the FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles and EU open science policies.

This document outlines the EuroFAANG infrastructures expectations for handling data generated within and in partnership with the Research Infrastructure, and the principles of accessing data generated as part of the infrastructure. This policy document should be viewed in conjunction with the EuroFAANG Guide for Transnational Access, that is also available from the EuroFAANG website (<https://eurofaang.eu/>).

By implementing this open data access policy, the EuroFAANG RI aims to contribute to a more transparent and collaborative research ecosystem advancing knowledge discovery, accelerating scientific progress, and ensuring exploitation of results, to enable effective genotype to phenotype research in farmed animals by the European research community and beyond.

## **1. The EuroFAANG RI**

The EuroFAANG RI builds upon the global Functional Annotation of Animal Genomes (FAANG) initiative to harness the power of data to illuminate the Genome to Phenome (G2P) link in farmed animals (both terrestrial and aquatic). By leveraging this genomic and phenotypic information, EuroFAANG aims to empower researchers and innovators to tackle the pressing challenges facing the agri-food sector: an expanding population, environmental shifts, and concerns about animal welfare. EuroFAANG's core mission is to foster research and innovation around G2P prediction in farmed animals. This data-driven approach seeks to revolutionize European animal production, health and welfare, making it sustainable, efficient, and aligned with societal values.

## **2. EuroFAANG Open access policy**

As an infrastructure, EuroFAANG will be supporting the community generation of open data and research output for genotype to phenotype research. Open access to EuroFAANG data is crucial both for verification of results and to ensure community reuse. The EuroFAANG RI promotes open access to and reuse of research data, following the "as open as possible, as closed as necessary" principle. By default, all data generated at the EuroFAANG RI will be openly shared unless legitimate reasons warrant a more restricted access.

The scope of legitimate restriction is being explored in detail by the EuroFAANG RI as part of its developing Transnational access (TNA) policy in work package 2. This policy outlines that the infrastructure in principle strives to be open, but the inclusion of industrial resources and data with Intellectual Property rights also offers clear advantages and

increases research and innovation around G2P prediction in farmed animals through fostering collaboration between industry and research communities. As the policy surrounding the transnational access to the RI by industrial users is further developed, the requirements for data access restrictions will also be put in place to provide the necessary and legitimate protection of Intellectual Property rights, whilst ensuring overall value for the wider research community and EU open science requirements of publicly funded research. The data policy for different access scenarios will be developed further based upon the types of use and types of users, as outlined in the Transnational Access Policy (for the latest version of this policy please see the EuroFAANG website (<https://eurofaang.eu/>)).

### ***Types of Use***

- A: Retrieve data stored at the infrastructure and use locally
- B: Obtain biological resources stored at the infrastructure and use locally
- C: Online access to and use of the compute resources of the infrastructure
- D: On site access to and use of the compute resources of the infrastructure
- E: On site access to and use of the experimental facilities of the infrastructure

### ***Types of Users***

- 1: Academic user fully open access
- 2: Academic user partial open access (e.g. for use of the infrastructure within an ongoing collaboration between an academic and industrial partner)
- 3: Industrial user fully open access
- 4: Industrial user partial open access
- 5: Industrial user, private use only

## **3. FAANG Data Sharing Policy**

The EuroFAANG RI will follow the global FAANG Data Sharing Policy. Fully open access users are required to follow the FAANG Data Sharing principles below, and all data consumers will need to follow the provisions against each dataset as recorded in the public archives and FAANG Data Portal.

Exceptions to this policy will be granted based on the types of use and types of users as defined by the Transnational Access (TNA) Policy. The process for granting this, for example to protect Intellectual Property rights for industry users of the RI, will be clarified and recorded here once these processes for TNA is further developed during the concept development phase of the EuroFAANG RI, as the RI establishes its processes for access and types of access. Substantial progress has already been made in Deliverables 2.1 and 2.2 and will continue to develop as stakeholder discussions continue.

For the latest version of this policy please see <https://www.faang.org/data-share-principle>, but for clarity the version as per the time of publishing of this document is recorded below.

# The FAANG Data Sharing Statement

**Version 2.0**  
(December 1, 2021)

## Definitions

**Archive** means one of the archives hosted at the EMBL-EBI, NCBI or DDBJ. These include the ENA, Genbank, ArrayExpress and Geo. A full list of the FAANG recommended archives is available as part of the FAANG metadata recommendations.

**Submission** means data and metadata submission to one of the FAANG recommended Archives.

**FAANG member** means an individual who has signed up to the FAANG consortium through the FAANG website and agreed to the FAANG core principles.

**Data** means any assay or metadata generated for or associated with FAANG experiments.

**Analysis** means any computational process where raw assay data is aligned, transformed or combined to produce a new product.

**Primary analysis** results consist of sample level analysis such as alignment to a reference genome or quantification of signal in the assay.

**Integrated analysis** results represent analyses which draw together data from multiple samples and/or experiments such as genome segmentation or differential analysis results.

**Internal** means data that is only accessible via the FAANG private shared storage.

**Private** shared storage means a storage space hosted at EMBL-EBI that has access limited to agreed persons by the data provider

**Public** means all data is available through the FAANG public data portal and underlying public archives, without embargo and is accessible to everyone.

This document describes the principles of data sharing for the FAANG consortium. Any queries about this document should be sent to [faang@iastate.edu](mailto:faang@iastate.edu) and [faang-dcc@ebi.ac.uk](mailto:faang-dcc@ebi.ac.uk).

FAANG believes that pre-publication data-sharing, collaboration and data reuse is for everyone's benefit and is strongly encouraged.

### **For FAANG data consumers:**



FAANG data are released under the Fort Lauderdale and Toronto principles 1,2. FAANG data creators reserve the right to first publication of the results obtained from using a dataset in genome wide analysis (see box 1 for clarifying examples). The publications made on any dataset can be checked on the FAANG Data Portal (<https://data.faang.org/>). If you are unsure if you are allowed to publish on a dataset, please contact the FAANG Data Coordination Centre and FAANG consortium (email [faang-dcc@ebi.ac.uk](mailto:faang-dcc@ebi.ac.uk) and cc [faang@iastate.edu](mailto:faang@iastate.edu) to enquire.)

When using FAANG data you should **cite relevant publications and preprints from the data creators as well as all of the data accession numbers (e.g. PRJEB19199) in the main body of the publication** (not in the supplementary materials).

The FAANG consortium is producing high quality and well-annotated datasets to support the community in generating a powerful genome to phenome resource and promotes rapid dissemination of data to accelerate research. FAANG datasets are high quality, focus on a standardised set of multi-omic assays, are accompanied by rich validated metadata, phenotypic information and detailed protocols. FAANG participants

### Box 1

Examples of **permitted use**, that must include **citation of relevant publications or preprints from the data creators and the dataset accession numbers** in the resulting manuscript:

Any researcher may download sequence data and/or derived bed files from the data portal, map these data to a genome and may derive results from these mapped data to address limited questions in their own research projects such as:

1. Is a specific set of genes expressed in a distinct tissue or set of tissues?
2. Is a locus, or pathway impacted by a particular histone mark?
3. Are particular SNV allele(s) present in the FAANG dataset?
4. What functional elements are present in a genomic region of interest for a particular trait?

Examples of **prohibited use** without prior publication from the data creators or **permission from the author**:

What is prohibited is the publication either on-line, or in the peer reviewed literature, of the results of a genome wide analysis of these data. Examples include but are not limited to:

1. Publishing on-line or in the peer reviewed literature a genome wide gene annotation file (gtf or bed) detailing transcription and isoform variation for the species' genome.
2. Publishing on-line or in the peer reviewed literature a genome wide survey of allele specific expression of transcripts and isoforms.
3. Publishing on-line or in the peer reviewed literature results derived from an integrated analysis of these data with other datasets for a genome wide study.

The above examples are not an exhaustive list, **if in doubt, please contact the FAANG Data Coordination Centre and FAANG consortium (email [faang-dcc@ebi.ac.uk](mailto:faang-dcc@ebi.ac.uk) and cc [faang-contact@animalgenome.org](mailto:faang-contact@animalgenome.org)).**

provide these data pre-publication to encourage data reuse for maximal benefit to the community.

The FAANG Steering Committee commits to report to journal editors and the laboratories involved **any event that disregards the rights of data creators** (including biological measurements as well as analysis of such measurements).

Fostering collaboration through joint data analyses is also highly encouraged so you are invited to contact data creators directly (or via [faang-dcc@ebi.ac.uk](mailto:faang-dcc@ebi.ac.uk)), or seek collaborative partners amongst the FAANG working groups and membership.

### **For FAANG data producers:**

FAANG recognizes that rapid sharing of the sample metadata and raw data generated by the consortium with the wider community is a priority. FAANG aims to ensure that everyone can benefit from the data created by FAANG to aid their own research as rapidly as is possible.

- All sample metadata and raw data produced for a FAANG associated project will be submitted to the public archives, without any hold until publication date, as soon as possible after sampling or data generation and initial quality control checks.
- All primary and integrated analysis results produced for a FAANG associated project are also encouraged to be made public prior to publication without embargo. However, it is acceptable that primary and integrated analysis results are kept private until publication, as long as the sample metadata and raw data have been made public.
- All FAANG public data are released under Fort Lauderdale and Toronto principles 1,2. The FAANG website, dataset descriptions and Data Portal have clear data reuse statements. The FAANG submission guidelines describe the suggested statement to include with your dataset submissions ([https://dcc-documentation.readthedocs.io/en/latest/experiment/ena\\_template/](https://dcc-documentation.readthedocs.io/en/latest/experiment/ena_template/)).
- The Data Portal has developed mechanisms to clearly identify which datasets are unpublished and which have at least one publication.

For FAANG primary and integrated analyses not made available in archives pre-publication, FAANG recognizes the need to enable and promote collaboration amongst consortium and community members. FAANG therefore provides functionality for primary and integrated analyses to be privately shared between FAANG members in private shared storage hosted at the EMBL-EBI. This requires an agreement between the two parties and that all have agreed to **the Fort Lauderdale and Toronto principles** <sup>1,2</sup>.

Only FAANG data can be submitted to the FAANG Data Portal.

All members of FAANG can and will continue to do experimental and analysis work outside of FAANG and the other data generated is not required to meet the same data sharing expectations.

Software and analysis pipelines developed by FAANG consortium members are strongly encouraged to be released under permissive open source software licenses wherever possible, such as Apache 2.0.

The FAANG Steering Committee commits to report to journal editors and the laboratories involved any event that disregards the rights of data creators (including biological measurements as well as analysis of such measurements).

#### REFERENCES:

1. [Fort Lauderdale principles: Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects.](#) (alt link)
2. [Toronto International Data Release Workshop: Rapid release of prepublication data has served the field of genomics well. Attendees at a workshop in Toronto recommend extending the practice to other biological data sets.](#)

#### Version

**2.0**

Update approved by the FAANG steering committee on 1st December 2021; [Original](#) approved on 26th May 2015.

## 4. FAIR data policy

The FAANG/EuroFAANG Data Coordination Centre and FAANG/EuroFAANG Data Portal will ensure that all data generated in the RI meets the highest possible FAIR principles.

This will ensure that the data will be:

- **Findable:** Data should be easy to find for both humans and machines. EuroFAANG RI will use consistent naming conventions, providing clear and accurate metadata, and registering data in discoverable public INSDC repositories.
- **Accessible:** Data should be accessible to anyone in the research community (noting that it is expected that some generated data from industry may carry additional restrictions to protect Intellectual Property). EuroFAANG RI will provide open access to data whenever possible, and ensure that users provide standard formats and mandatory protocols for data access.
- **Interoperable:** Data should be interoperable with other data sets. EuroFAANG RI will ensure the use of a common FAANG data model and ontologies, and providing documentation that describes the data in a way that others can easily understand.
- **Reusable:** Data should be reusable for other purposes than the original study. This means ensuring clear information about the provenance of the data, extensive

rich metadata, and open licenses that allow others to reuse the data whenever possible.

To achieve this, users of the EuroFAANG RI must comply with the below provisions laid out for Brokered submission requirements and metadata standards and Data access procedures.

## **5. Long term data management and preservation of data**

The EuroFAANG RI is committed to long-term data preservation and stewardship. Data will be properly documented, archived, and backed up to ensure its integrity and future accessibility. All data generated within the EuroFAANG infrastructure will have internationally recognised identifiers of the International Nucleotide Sequence Database Collaboration (INSDC). Submission through the EuroFAANG/FAANG data portal ensures that data is publicly archived in an INSDC archive. These will be issued upon submission to the EMBL-EBI BioSamples and European Nucleotide Archives. These archives ensure long term preservation and assurance of data beyond the availability of any community specific portals and data services.

## **6. Brokered submission requirements and metadata standards**

All EuroFAANG data will be submitted through the EuroFAANG/FAANG brokered submission system that will ensure compliance with EuroFAANG's metadata standards and the data format standards that it develops in collaboration with Elixir, key EU research infrastructures in animal science, and the animal phenotyping community represented in EU infrastructure projects. The data submission system and detailed instructions are available here <https://data.faang.org/validation/samples>. Further collaboration with Elixir is expected to take place in the frame of the proposed focus group 'Domestic Animal genomes and phenomes' (see D7.1).

All users of the EuroFAANG RI will be required to meet the minimum EuroFAANG/FAANG metadata standards and encouraged to submit metadata as richly as possible. The brokered submission system validates and ensures the minimum standards are met and also suggests further improvements to enrich metadata reporting. This provides a submission process for samples, raw datasets and analysed datasets, with ontologies required to be supplied for many fields. Submissions also require mandatory detailed protocols and links to data workflows used to generate data.

Detailed instructions on making a EuroFAANG submission are provided here <https://dcc-documentation.readthedocs.io/en/latest/>

## **7. Access procedures**

Data access will be facilitated through the dedicated EuroFAANG/FAANG data portal or directly from the INSDC public archives. Researchers accessing data need to adhere to the terms and conditions of data use, including citation requirements and limitations on redistribution. In the case that EuroFAANG will coordinate more restricted datasets, pertaining to the outcomes of TNA design, the EuroFAANG RI will evaluate data access requests promptly and fairly, considering the potential benefits and risks of sharing. This requires a close collaboration in the development of TNA (Work package 2, WU) and Data policies (Work package 3, EMBL). Applicants will be notified of the decision and provided with instructions for accessing the data if their request is approved.

## **8. Data Re-use and Citation**

Users are encouraged to reuse data from the EuroFAANG RI for further research and innovation. Proper citation of the data is mandatory to acknowledge the data creators and contribute to responsible research practices. Data derived from the use of the infrastructure that is submitted to public repositories should acknowledge the EuroFAANG RI and include the INSDC project identifier in the main text of the publication. Publications, be it in peer reviewed journals, presentations at conferences or in papers directed to the public, should also acknowledge the use of the EuroFAANG RI.

## **9. Role of EuroFAANG Data Coordination Centre**

Data will be collated in the context of existing EuroFAANG, FAANG and community datasets in the FAANG Data Portal. The existing EuroFAANG projects and EMBL Data Coordination Centre pride themselves on ensuring the reusability of generated data and research outputs by providing rich supporting metadata, detailed mandatory protocols of research and analysis methods, links to the open access analysis software and parameters that generated the data, and clear provenance and licensing. The EMBL-EBI EuroFAANG/FAANG Data Coordination Centre maintains the data standards, submission infrastructure and FAANG Data Portal.

## **10. Monitoring and Review**

This policy will be reviewed and updated periodically to reflect evolving open science practices, EuroFAANG RI data management and legal frameworks. User feedback and input are welcome to ensure this policy remains relevant and effective. The policy is under active development whilst the RI is established, and it is expected that many sections will be updated based on the recommendations of EuroFAANG RI think tanks and workshop events.

## **11. Contact and Support**

For any questions regarding data access, data management, or this policy, please contact the EMBL EuroFAANG RI Data Coordination Centre at [faang-dcc@ebi.ac.uk](mailto:faang-dcc@ebi.ac.uk).

## 12. Additional Considerations

- Please note that this data policy may be supplemented by specific data use clauses for complex or sensitive data as part of TNA agreements between the EuroFAANG RI and users. Industry users in particular should contact the EuroFAANG RI to discuss specific data access and data generation requirements.
- In case it is anticipated that the use of the EuroFAANG infrastructure could result in data, resources or procedures that can be protected with IP rights, the use and ownership of such IP rights need to be agreed before commencing the TNA. Any background knowledge (e.g. information, know-how, data, or material, including any IP Rights pertaining to such knowledge) that is held by the user prior to use of the EuroFAANG RI also needs to be described in specific clauses in the TNA agreement before commencing with the TNA.
- The EuroFAANG RI may collaborate with other research infrastructures and data repositories to facilitate data sharing and interoperability.
- User training and resources will be available to support researchers in effectively accessing and reusing data from the EuroFAANG RI, please consult the EuroFAANG website for further information (<https://eurofaang.eu/>).
- This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101094718. The content of this document reflects only the consortium's view. The European Commission is not responsible for any use that may be made of the information it contains.

### 3. Conclusions

The above data policy and access principles document, represents the first version of EuroFAANG's commitment to open science principles and the role it will play in effective genotype to phenotype research in farmed animals by the European research community and beyond. The policy has outlined the principles and processes of the interoperability of the different infrastructure elements that facilitate data coordination and management for the RI, through the connection to the Data Coordination Centre at EMBL, transfer and curation of data in the INSDC public archives and presentation of data through EuroFAANG (FAANG) data portals and related services. It is expected that the policy will continue to evolve as the RI develops, to further outline the data handling expectations and data access requirements for all users of the EuroFAANG infrastructure, with a focus on open science, FAIR data principles and existing European policies. This policy will cover the interoperability of the different infrastructure elements, connection to the Data Coordination Centre at EMBL, transfer and curation of data in the public archives and presentation of data through FAANG data portals and related services. Work package 2 and 3 will work closely with industry stakeholders to establish effective TNA data access and sharing clauses for EuroFAANG's TNA and Data Access policies. The Data Policy and Access Principles will be published on the EuroFAANG website (<https://eurofaang.eu/>) alongside the formalised Transnational Access Policy for the infrastructure.